

# Animal Experimentation and Its Relevance to Man

by D. G. Hoel\*

The problem of quantitatively estimating human cancer risk based upon animal carcinogenesis studies is reviewed. Mathematical functions for dose-response relationships are discussed with particular emphasis on multistage models. These models are based upon a single cell somatic mutation theory for the carcinogenesis process. It is shown that the multistage model and others which incorporate background additively are well approximated in the low dose region by a linear function. The relationship between time-to-tumor and the multistage model is indicated. This relationship is important when dealing with less than life time exposure such as with data from many occupational studies. Design of bioassay experiments and its impact on risk estimation is noted. Finally, the problem of species-to-species extrapolation is considered.

## Introduction

The problem of identifying human carcinogens and assessing their potency will often depend, out of necessity, on laboratory experimentation. This ranges from DNA chemistry to chronic exposures in lifetime animal studies. Epidemiological studies in carcinogenesis are naturally more appropriate for assessing human risk. Unfortunately they require long term prior exposures to man which do not exist with new compounds. For older compounds laboratory and epidemiological studies hopefully interact in a creative manner.

The problem I wish to discuss is how one can estimate human risk to a carcinogen based upon laboratory studies. This question involves several problems concerning the appropriateness of animal models for predicting effects in man. In the standard lifetime, rodent bioassay such as the one used by the National Cancer Institute there are several ways in which to arrive at incorrect conclusions. For example, often concerns are expressed about the statistical errors relating to sample size, etc., and also to the possible biological errors in the sense that the particularly chosen rodent strains may not be

relevant to man for the type of carcinogen under consideration.

If a particular animal model is a reasonable qualitative predictor of human cancer, how reasonable is it as a quantitative predictor of carcinogenesis potency in man? Clearly, if the regulation of environmental carcinogens involves risk-benefit considerations, we must have acceptable methods for estimating effects in man at low exposure levels based upon laboratory animal studies at relatively high exposure levels. These high experimental levels are necessitated by limited size (cost) of the animal study and the need for observed experimental effects. This raises interesting experimental design considerations. On the one hand we require high statistical power for confident qualitative identification of carcinogenesis, and this is obtained at high dose levels. On the other hand experimentation at low dose levels is desired for predicting effects at low exposure levels.

## Dose-Response Models in Carcinogenesis

In order to estimate carcinogenic effects, some assumptions or models are needed which relate the frequency of tumor with exposure level of the carcinogen in question. If the true form of the dose-response function were known, one could correctly

---

\*National Institute of Environmental Health Sciences, National Institutes of Health, P.O. Box 12233, Research Triangle Park, North Carolina 27709.

estimate the effects of the carcinogen at low levels within a stated degree of statistical uncertainty. This would provide the first step in a human risk assessment. The low dose estimates based upon the experimental animal data and dose-response model must then be extrapolated to man. This requires a considerable number of biological assumptions and is an uncertain process at best.

Several quantitative theories of carcinogenesis have been proposed which relate tumor incidence with both dose rate and duration of exposure of the carcinogen. These models generally assume that the carcinogenesis process is single cell in origin and is the result of several stages which can include somatic mutation. The transitional events are individually assumed to depend linearly on dose rate. This then leads in general to a model in which the probability of tumor is approximately a low order polynomial in dose rate. In the low dose region which would relate to environmental levels, one finds that the response is well approximated by a linear function of dose rate (1-4). Another class of dose-response models which have been applied are those which have a history in bioassay work. For example, both the logistic and probit functions are used for estimating LD<sub>50</sub> in toxicology. Mantel and Bryan (5) have applied the probit to low dose estimation in carcinogenesis. Its drawback is primarily that the model is not based upon a mathematical description of a biologic mechanism as is the case with the multistage somatic mutation model. Since it produces quite different results in the low dose region it probably should not be applied in those cases where it is believed that the carcinogen is direct acting with cellular DNA as is proposed in the multistage models (4, 6).

Of the multistage models, the one discussed by Armitage and Doll (3) has received considerable attention (4, 6-9). The function which is used states that the probability of a tumor by time  $t$  using dose rate  $d$  is

$$P(d, t) = 1 - \exp \{-g(d)h(t)\}$$

where

$$g(d) = \prod_{i=1}^k (\alpha_i + \beta_i d)$$

with  $\alpha_i \geq 0, \beta_i \geq 0$  for  $i = 1, \dots, k$ . This represents a  $k$ -stage or mutational step model with dosage and time factoring. The above mentioned authors have developed maximum likelihood estimates for the unknown parameters in the model. Further they have approximate confidence estimates on the low dose estimates of the probability of tumor. These procedures are mathematically complicated in that com-

puting algorithms are required in order to obtain the estimates. Recently Guess et al. (6) applied these algorithms to experimental carcinogenesis data on vinyl chloride, DDT, chloroform, DMN, and diel-drin. In all of these cases  $k = 1$  or  $2$  was found to provide the best fit, although DMN yielded a zero estimate of the linear coefficient. In each case the upper bound of risk was estimated to be linear at low dose as one would theoretically anticipate.

It was previously mentioned that the multistage models were approximately linear at low dose rates. An important question is: how good is the linear approximation? It may be that linear extrapolation is a perfectly adequate method and much simpler than using complicated computer algorithms. Furthermore, there is the intuitive belief that the lower convex portion of a typical dose-response curve would be bounded above by a straight line (10). It is felt, however, that a line is a too conservative upper bound in that it greatly overestimates the risk at low dose levels. It has been shown by Crump et al. (4) that the linear approximation to the multistage model is quite good if a background or spontaneous tumor rate is present. For example, we define  $d_l$  and  $d_m$  as the dose associated with a given small probability of tumor in excess of background when we use a linear model and a multistage model, respectively. Let  $k$  denote the number of stages in the carcinogenic process and define  $p$  by

$$(1 + p)P(0) = P(d_e)$$

where  $d_e$  is the experimental dose and  $P(d)$  is the probability of tumor at dose  $d$ . It was then shown by Crump et al. (4) that approximately

$$d_m/d_l \leq p/\{k[(1 + p)^{1/k} - 1]\}$$

Table 1. Upper bound values of  $d_m/d_l$  for selected  $k$  and  $p$ .

$k$	$d_m/d_l$				
	$p = 0.1$	$p = 0.5$	$p = 1$	$p = 10$	$p = 100$
1	1	1	1	1	1
2	1.02	1.11	1.21	2.16	5.52
5	1.04	1.18	1.35	3.25	13.18
$\infty$	1.05	1.23	1.44	4.17	21.67

Thus linear extrapolation in the lower convex portion of the dose response curve is both a reasonable upper bound and direct estimate of the probability of tumor (Table 1). This is an appropriate procedure to use unless there is specific information concerning the mechanism of the carcinogenic process which would indicate a model other than the multistage. This might be the case, for example, with certain promoters or cocarcinogens.

The models discussed so far have been very simplistic in nature. The multistage model assumes that the mutational step is proportional to the administered dose rate. This ignores the influences of pharmacokinetics which may indicate a changing proportion of active metabolite with administered dose level. Also immune response and DNA repair systems are not taken into account. Even without these biological complications the models are fairly complex, which indicates that we are still a long way from a nonlinear methodology in which one would have much biological confidence.

## Time To Tumor

Most dose-response studies in animal experimentation are plots of dose rate versus total tumor incidence for the duration of the study. In large studies there is often sufficient data to plot cumulative tumor incidence versus age. This is then done for each dose and the curves compared. Models are also fit to data and have the possibility of use in low dose estimation. Albert and Altshuler (11) have fit the log-normal model of Blum and Druckery to several sets of laboratory data. Also Whittemore and Altshuler (12) have fit both the Weibull and log-normal model to the physician smoking data of Hill and Doll. As with the ordinary dose-response models one finds that there usually are not sufficient data to distinguish between competing models. Yet the choice of a time-to-tumor model can greatly change any estimated low dose effects.

It has been suggested (13) that the time to first tumor or latency period increases with decreasing dose. Then with a sufficiently small dose the time to first tumor will exceed man's lifetime, thus producing an effective carcinogenic threshold. Unfortunately as shown by Guess and Hoel (14), the observed time to first tumor will increase with decreasing dose for the multistage model, yet no true threshold will occur at low doses with this model. In other words, a non-threshold model will predict that the data will look as though a threshold in time exists.

Using the Armitage and Doll (3) multistage model with the incorporation of time-to-tumor one has for the age-specific incidence rate

$$I(t, d) = kt^{k-1} \prod_{i=1}^k (\alpha_i + \beta_i d)$$

with  $k$  the number of stages and  $d$  the dose rate. The power of the duration of exposure  $t$  corresponds to what is seen with vital statistics data on human cancer types. Whittemore and Altshuler (12) found, for example, with the lung cancer and cigarette smoking data that  $I(t, d) = ct^5d$ , where  $t$  represented

the duration of smoking while  $d$  was the daily consumption rate. This shows the importance of the duration of exposure in any attempt to estimate risks.

## Design and Detection

The routine screening of compounds for carcinogenic activity is a slow and expensive process using lifetime rodent studies. These assays, such as used by NCI, are designed for detection of activity and are not intended to provide dose response information needed for low dose risk estimation. It is necessary to use high dose experiments in order to generate as much statistical power as possible. Even so, weak carcinogenic activity may very well go undetected. For example, suppose one is comparing 50 control animals with 100 treated animals using Fisher's exact test and nonrandomized decisions. Then for  $\alpha = 0.01$  (one-tailed) one has less than a 5% chance of detecting an increased tumor incidence of 5% over the background rate. This would essentially go unnoticed. In actual assays both sexes and maybe two species will be used. This will increase the power, assuming an appropriate decision rule is applied which considers the multiple testing of many tissues besides the replications in sex and species. Fears et al. (15) have looked into these issues in depth.

If we want to estimate effects at low dose levels, the high dose bioassay often will be quite useless. For example, if the majority of the experimental animals are positives at the lowest experimental dose, then little can be said about when the dose-response function turns over. In this case a straight line may not provide an upper bound to the curve. Crump and Langley, in an unpublished study have considered the simplest case of no background and a single experimental dose at  $d$ . Suppose we require an upper bound  $\hat{d}_a$  to a dose  $d_a$  which corresponds to the probability of cancer  $P_a = P(d_a)$  and is assumed to be small. We define the upper bound  $\hat{d}_a$  by

$$P\{\hat{d}_a \leq d_a\} \geq 1 - \alpha$$

and use

$$\hat{d}_a \equiv d_t P_a / \hat{P}_t$$

where

$$P\{\hat{P}_t \geq P(d_t)\} = 1 - \alpha$$

Now  $\hat{d}_a$  corresponds to an acceptable dose, and thus we can select the experimental dose  $d_t$  in order to maximize the expected value of  $\hat{d}_a$ . Crump and Langley showed that

$$E(\hat{d}_a)P_a^{-1} = d \sum_{r=0}^N g(r)^{-1} \left(\frac{N}{r}\right) \Theta^r(d_i) \{1 - (d_i)\}^{N-r}$$

where  $\Theta(d)$  is the true dose-response function,  $N$  the experimental sample size at dose  $d_i$ , and  $g(r)$  in the upper bound  $\hat{P}_i$  based on  $r$  tumors. Using this expression, Crump and Langley found considerable effects on  $E(\hat{d}_a)$  by the choice of  $d$ . For several functional forms of  $\Theta(d)$  it appeared as if  $d_i$  is optimized by using approximately one-half the dose which produces a 10% tumor incidence. This should be a useful rule of thumb for single dose experiments.

More recently Hoel and Jennrich (16) have obtained optimal designs for the multistage model using techniques of Chebyshev systems in polynomial regression. If

$$P(d) = \{1 - \exp - \sum_{i=0}^k \beta_i d^i\}$$

and  $k + 1$  design points are used they obtain the design which minimizes  $\text{Var}\{\hat{P}(d_0) - \hat{P}(0)\}$  where  $d_0$  is the dose of interest and  $\hat{P}(d_0)$  the estimated incidence at  $d_0$ . They found that their optimal design decreased the variance by a factor of 3 when applied to Maltoni's 1975 vinyl chloride study (17). Maltoni placed his low dose experimental points equally at  $d = 0, 50, 250$ , and  $500$ . The Hoel and Jennrich design suggests 27% at  $d = 0$ , 53% at  $d = 83$ , 15% at  $d = 342$ , and 5% at  $d = 500$ . The design question, which is still open, is how would one combine detection with low dose estimation. The problem is to have both high dose levels for power of detection and also low dose levels for risk estimation.

## Species-to-Species Extrapolation

The problem of the appropriateness of the rodent carcinogenesis study for human risk prediction is most difficult. One usually begins by stating that of the 25 known human chemical carcinogens all except for possibly arsenic and benzene are also rodent carcinogens. There is good qualitative agreement between laboratory results and epidemiological studies. There are, however, few well done human studies as compared to the hundreds of compounds studied in rodent bioassays. Tomatis (18) discussed the predictiveness of long-term rodent studies. He points out the early identification of the carcinogenicity of diethylstilbestrol, 4-aminobiphenyl, and vinyl chloride through experimental animal testing.

Recently short-term *in vitro* testing has been compared with the long-term animal studies. Tests such as Ames' Salmonella mutagenicity assay are felt to be presumptive of carcinogenicity. This is based upon the proposition that many types of cancer are

the result of somatic mutations and often are single cell in origin. Recently an IARC working group (19) expressed the opinion that

"When there are inadequate animal data, positive results in validated short-term tests are an indication that the compound is a potential carcinogen and should be tested in animals for an assessment of its carcinogenicity. Negative results in short-term tests cannot be considered sufficient to rule out carcinogenicity."

By the Ames test (20, 21) most of the chemical carcinogens tested so far are mutagenic (157/175) and most noncarcinogens are not mutagenic (95/108). Further, there is some preliminary work suggesting that cancer potency in animal studies is quantitatively predicted by the Salmonella test (22). Much more work is needed in this area, since few compounds have been compared in this manner.

Meselson has considered in a NAS report (23) several human carcinogens (benzidine, chlor-naphazine, DES, aflatoxin B<sub>1</sub>, vinyl chloride, and cigarette smoke) and compared the sensitivity of man with laboratory animal species. He found using the most sensitive animal species as a predictor for man to be approximately correct for benzidine, chlornaphazine and cigarette smoke. For the other three compounds man was observed not to be as sensitive as predicted from the animal studies.

Studies such as the one conducted by Meselson will, at least empirically, answer the question as of the precision of risk estimation. The process of quantitatively estimating carcinogenic effects in man will improve especially with better incorporation of pharmacokinetic considerations and understanding of environmental and genetic differences in human populations. With respect to the precision of risk estimates, however, there is not an apparent theoretical method for quantifying biological errors. Thus we are dependent upon empirical experience which is somewhat inadequate at the present time.

Rall (24) has discussed species differences in carcinogenesis testing in some detail, paying particular attention to pharmacological differences and similarities in the species. He concludes that "laboratory animal carcinogenicity tests predict well for man and that such tests do offer a mechanism by which the prediction of human carcinogenesis is possible before human exposure and with reasonable accuracy."

In gathering empirical evidence for the qualitative predictability of animal studies for man, careful use of epidemiological evidence is needed. The most straightforward situation is when one deals with data from direct studies. Lung cancer and smoking is a good example where careful modelling of both level

and duration of exposure has been considered (25, 12). From this study one feels that accurate estimates of risk can be made, at least for British physicians who smoke. Extending these estimates to the general population may still be quite reasonable. This assumes that we are interested in the median man and feel that the British physician is an adequate model. Errors of magnitude do occur, however, when we attempt to estimate the risks to susceptible subgroups in the population such as uranium miners and asbestos workers. Genetic and environmental susceptibles are often not identified nor predicted. Thus synergistic effects are a real problem and often will be unanticipated.

Ecological or indirect studies are much more difficult to assess with respect to both predictability and precision. These studies attempt to correlate vital statistics data and environmental exposure data. Further, the studies attempt to quantitatively predict health effects based upon exposure data usually employing simple regression models. Examples include mortality rates and various air pollutants. Also chloroform levels in drinking water and bladder, colon, and rectal cancer rates have been studied recently. It is known that the predicted health effect often will vary depending upon the statistical method of analysis used. The reliability and thus usefulness of this quantitative approach to risk estimation is still not understood.

Estimates are needed for risks to the general population from various environmental agents when the only human data available is from high exposure groups. For example, risk calculations have been made for populations living in the vicinity of vinyl chloride plants. These estimates are extrapolations from highly exposed worker populations.

Even with the application of dose response and exposure duration models, childhood and *in utero* exposures are not included in the model. Thus serious biological errors could be made. It is the judgement of the possibility of such model errors which is needed for the quantification of the precision of the risk estimates. The statistical aspects can be properly assessed. What is needed is quantification concerning the likelihood of the model being representative of the true state of nature.

Finally, one technical point should be discussed. In the comparison of human data with animal data we are often dealing with less than lifetime human exposures. Much of the human data is obtained from industrial exposures. Applying the multistage model one finds a critical dependency on the duration of exposure. As discussed earlier the age specific rate often is of the form  $dt^{k-1}$  where the number of stages  $k$  is 4, 5, or 6. Now, for a human study with 20 years

exposure and followup one would underestimate the life time risk by a factor of about 500  $[(70/20)]^{k-1}$  with  $k = 6$ , and it is not taken into account. Thus we see the care with which the human data must be treated when quantitative comparisons to animal studies are made.

## Conclusion

The research area of low dose risk estimation has and will continue to be of statistical interest. Dose-response modelling will also continue although we must be on guard not to lose sight of the assumptions made. There seems always to be the danger of providing more mathematical sophistication than the biological knowledge warrants.

Statistical errors or confidence statements do not include a quantification of biological errors (e.g., wrong mathematical or biological model). Since the biological errors cannot be easily if at all quantified they tend to be ignored. The possible biological errors could easily overwhelm the statistical errors and thus considerable care is needed in the interpretation of confidence statements.

Specific research needs from a statistical standpoint include the following: (1) additional bioassay design of experiments research, including improved methods of statistical testing as well as risk estimation; (2) incorporation of pharmacokinetic considerations including repair systems in dose-response estimation; (3) empirical studies which attempt to quantitate the effects in man with those in laboratory animal experiments; (4) attention to quantifying the effects of susceptible subgroups and also synergistic activity.

## REFERENCES

1. Iversen, S., and Arley, N. On the mechanism of experimental carcinogenesis. *Acta Pathol. Microbiol. Scand.* 27: 773 (1950).
2. Nordling, C. O. A new theory on the cancer inducing mechanism. *Brit. J. Cancer* 7: 68 (1953).
3. Armitage, P., and Doll, R. Stochastic models for carcinogenesis. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. LeCam and J. Neyman. Eds., Univ. of California Press, Berkeley, California, 1961, pp. 19-38.
4. Crump, K. S., Hoel, D. G., Langley, C. H., and Peto, R. Fundamental carcinogenic processes and their implications for low dose risk assessment. *Cancer Res.* 36: 2973 (1976).
5. Mantel, N., and Bryan, W. R. Safety testing of carcinogenic agents. *J. Natl. Cancer Inst.* 27: 455 (1961).
6. Guess, H. A., Crump, K. S., and Peto, R. Uncertainty estimates for low-dose rate extrapolations of animal carcinogenicity data. *Cancer Res.* 37: 3475 (1977).
7. Crump, K. S., Guess, H. A., and Deal, K. L. Confidence intervals and test of hypotheses concerning dose response relations inferred from animal carcinogenicity data. *Biometrics* 33: 437 (1977).

8. Guess, H. A., and Crump, K. S. Low-dose rate extrapolation of data from animal carcinogenicity experiments — analysis of a new statistical technique. *Math. Biosciences* 32: 15 (1976).
9. Hartley, H. O., and Sielkin, R. L. Estimation of "safe doses" in carcinogenic experiments. *Biometrics* 33: 1 (1977).
10. Hoel, D. G., Gaylor, D. W., Kirschstein, R. L., Saffiotti, U., and Schneiderman, M. A. Estimation of risks of irreversible, delayed toxicity. *J. Toxicol. Environ. Health* 1: 133 (1975).
11. Albert, R. E., and Altshuler, B. Considerations relating to the formulation of limits for unavoidable population exposures to environmental carcinogens. In: *Radionuclide Carcinogenesis*, J. E. Ballou, Ed., AEC Symposium Series, CONF-72050, NTIS, Springfield, Virginia, 1973, pp. 233-253.
12. Whittemore, A., and Altshuler, B. Lung cancer incidence in cigarette smokers: further analysis of Doll and Hill's data for British physicians. *Biometrics* 32: 805 (1976).
13. Jones, H. B., and Grendon, A. Environmental factors in the origin of cancer and estimation of the possible hazard to man. *Food Cosmet. Toxicol.* 13: 251 (1975).
14. Guess, H. A., and Hoel, D. G. The effect of dose on cancer latency period. *J. Environ. Pathol. Toxicol.* 1: 279 (1977).
15. Fears, T. R., Tarone, R. E., and Chu, K. C. False-positive and false-negative rates for carcinogenicity screens. *Cancer Res.* 37: 1941 (1977).
16. Hoel, P. G., and Jennrich, R. I. Optimal designs for certain cancer problems. *Biometrika* (submitted).
17. Maltoni, C. The value of predictive experimental environmental carcinogenesis — an example: vinyl chloride. *Ambio* 4(1): 18 (1975).
18. Tomatis, L. The value of long-term testing for the implementation of primary prevention. In: *Origins of Human Cancer*, H. H. Hiatt, J. D. Watson, and J. A. Winsten, Eds., Cold Spring Harbor Laboratory, New York, 1977, pp. 1339-1358.
19. IARC Internal Technical Report No. 77/002, (1977).
20. McCann, J., and Ames, B. N. Detection of carcinogens as mutagens in the *Salmonella*/microsome test: Assay of 300 chemicals: discussion. *Proc. Natl. Acad. Sci. (U.S.)* 73: 950 (1976).
21. McCann, J., and Ames, B. N. The *Salmonella*/microsome mutagenicity test: Predictive value for animal carcinogenicity. In: *Origins of Human Cancer*, H. H. Hiatt, J. D. Watson, and J. A. Winsten, Eds., Cold Spring Harbor Laboratory, New York, 1977, pp. 1431-1450.
22. Meselson, M., and Russell, K. Comparisons of carcinogenic and mutagenic potency. In: *Origins of Human Cancer*, H. H. Hiatt, J. D. Watson, and J. A. Winsten, Eds., Cold Spring Harbor Laboratory, New York, 1977, pp. 1473-1481.
23. National Academy of Sciences, National Research Council Report. *Pest Control: An Assessment of Present and Alternative Technologies*, Vol. 1, Contemporary Pest Control Practices and Prospects. The Report of the Executive Committee, Washington, D. C., 1975.
24. Rall, D. P. Species differences in carcinogenesis testing. In: *Origins of Human Cancer*, H. H. Hiatt, J. D. Watson, and J. A. Winsten, Eds., Cold Spring Harbor Laboratory, New York, 1977, pp. 1383-1390.
25. Doll, R. The age distribution of cancer: implication for models of carcinogenesis. *J. Roy. Statist. Soc. A* 134: 133 (1971).